A RESEARCHER'S GUIDE

to MEDI-CAL DATA

rev. November 2005

Prepared by

Jim Klein

Medical Care Statistics Section
California Department of Health Services
1501 Capitol Ave., Ste 6069
Sacramento, CA 95814
Telephone: (916) 552-8562
Email: jklein@dhs.ca.gov

ARNOLD SCHWARZENEGGER
Governor
State of California

Kim Belshé
Secretary
California Health and Human Services Agency

Sandra Shewry
Director
Department of Health Services

TABLE OF CONTENTS

APPENDICES

Introduction

This guide is intended to assist those performing research using Medi-Cal fee-for-service paid claims and managed care encounter data. For more information about Medi-Cal, researchers should review information provided on the Medical Care Statistics Section website, at http://www.dhs.ca.gov/mcss . The Manuals for the claims and encounter data referred to in this guide are available upon request.

1. Program Background

Medi-Cal (Medicaid in California) can take on different definitions, depending on the perspective of the researcher. The responsibilities of the Medical Care Statistics Section (MCSS) include ad hoc processing of medical Medi-Cal fee-for-service claims and eligibility data, and release of fee-for-service and managed care encounter data to the public. This Section is usually not involved in researching managed care encounter data, or claims for such non-Medi-Cal programs as the Child's Health and Disability Prevention (CHDP) program, the California Children Services/Genetically Handicapped Persons (CCS/GHPP) program, and the County Medical Services Program (CMSP). Note, however, that persons can be enrolled in both Medi-Cal and CHDP and/or CCS/GHPP, and thus their Medi-Cal claims could be accessed for reports by this Section.

For a more detailed description of Medi-Cal and the sources for service and eligibility data for Medi-Cal and non-Medi-Cal programs, please refer to Appendix A, Medi-Cal Description and Expenditures/Eligibles Sources.

2. Release of Data

2.1 Confidentiality

Although most data MCSS accesses can be made public, certain data must be kept confidential, pursuant to the California Public Records Act (Gov. Code, §§ 6250-6260), the Information Practices Act (Civ. Code, §§ 1798-1798.44), the California Welfare and Institutions Code (§§ 10850 and 14100.2), and, for Medi-Cal, regulations promulgated by the Federal Department of Health and Human Services . Beneficiary names and identification numbers normally cannot be released, although they can be encrypted in a consistent fashion so outside researchers can track the same person throughout their data sets.

Certain provider data are also confidential. The IRS identification number of Medi-Cal providers cannot be released, and the Medi-Cal provider is kept confidential on the advice of DHS Legal Services. Because the three-byte prefix, and the last byte, of the nine-character Medi-Cal provider

number carries meaningful information, all but these four bytes are encrypted. Provider names, addresses and earnings are public information.

A third area of confidentiality concerns rates negotiated by the California Medical Assistance Commission (CMAC) with Medi-Cal hospitals; these rates are confidential for four years. Generally, per diem rates could be determined using both the number of inpatient days and total monetary reimbursements made to these hospitals. Either of two following methods are thus used to protect these rates: 1) Total inpatient days or reimbursement amounts for these facilities are suppressed; or, 2) Because one hospital may sometimes have a CMAC contract which covers only some of their Medi-Cal services, inpatient claims data under CMAC and non-CMAC provider numbers can be combined before their release. The first method is usually the one used by MCSS, based on preferences by researchers.

Finally, the Health Insurance Portability and Accountability Act (HIPAA) privacy rules, implemented on April 14, 2003, have added new restrictions as to the type of entities that may receive personal medical data and for what purposes the data may be applied. For an overview of these, visit the MCSS website at http://www.dhs.ca.gov/mcss/GeneralInfo/HIPPA%20Privacy%20Rule.htm.

2.2 Data Media

Due to the extensive volume and complexity of Medi-Cal data, along with HIPAA restrictions, claims data are never released in their original file format, such as the 35-File. Claims and eligibility files in their native format are best run by staff with a Medi-Cal background, to ensure accurate translation of data and value formats. When the number of claims to be analyzed by the user is small, either the original claims files can be converted to simpler file formats (e.g., packed decimal fields converted to numeric, variable block records converted to fixed block) or to CD-ROM or DVDs media. MCSS can also provide the data as SAS data sets.

3. Eligibility

3.1 Eligibility Files

Eligibility for Medi-Cal is established at the county social welfare offices through on-line access to the State Medi-Cal Eligibility Data base maintained at the Health and Human Services Data Center. In addition, SSI recipients (aid codes 10, 20 and 60) have their eligibility established by the Federal Social Security Administration (SSA), makingf them automatically eligible for Medi-Cal. The eligibility data base also includes food stamp eligibility. Each record represents an individual currently or at some time in the last fifteen months eligible for Medi-Cal or food stamps. The on-line (VSAM) data base is used to create a Monthly Medi-Cal Eligibility File (MMEF) about the twenty-fourth of each month; this file is a shortened, snap-shot version of the on-line data base used to establish that month's eligibility along with retroactive eligibility information covering the previous

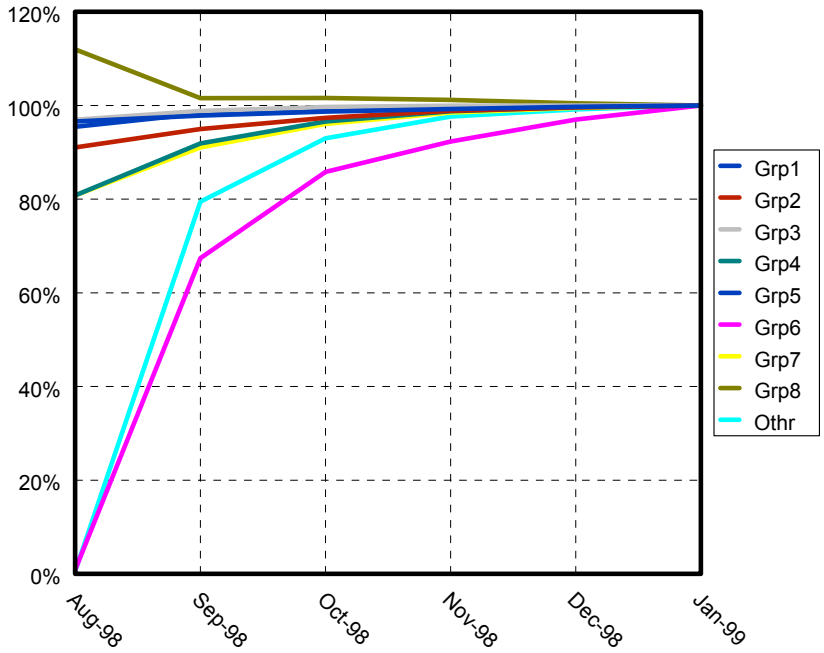
fifteen months.

As briefly discussed below in Section 3.4, Tracking Beneficiaries Across Claims, the counties issue a 14-digit beneficiary identification number consisting of the county code, aid code, a seven-digit family case number, a one-digit family budget unit, and two-digit person number. The SSA issues only the SSN, which is incorporated in this 14-digit ID as the last nine-digits. The family case number is, by and large, a reliable unique ID for tracking common households within a given county. A split in the household would result in a change to the FBU number. The convention the counties use for assigning the person number varies by county.

3.2 Querying Eligibility Files

Beneficiaries usually are eligible for Medi-Cal before the eligible month. However, a significant percent of all eligibles become eligible retroactively, for instance, due to meeting share of cost requirements and late Medi-Cal applications (e.g., after medical costs are incurred); aid categories where the latter is particularly the case are the medically needy and medically indigent. Chart 1, below, illustrates the increase of Medi-Cal eligibility counts for the month of eligibility August 1998 using eligibility files created August 1998 through January 1999. (See Appendix C for the table of numbers that were used to construct this chart.)

Chart 1, Increase of Percent of Eligibles by Aid Category



where:

Group 1 = SSI/SSP
Group 2 = Long Term Care
Group 3 = CalWorks - Cash Grant
Group 4 = Medi-Cal Only, Families
Group 5 = Medi-Cal Only, Aged, Blind, Disabled, no share of cost
Group 6 = Share of Cost
Group 7 = OBRA
Group 8 = Miscellaneous

This effect of significant growth in eligibility counts after a given month of eligibility requires the researcher to compromise between currentness and accuracy in obtaining counts: the more current, the less accurate, and vice-versa. In capturing data from past months' eligibility files, one data element sometimes used is the beneficiary address, including zip code. Zip codes are not kept on each monthly eligibility segment, but rather are overlaid with any new zip code for that month's eligibility file. Therefore, when using a file four months removed from the month of eligibility (MOE), the zip code of a beneficiary may have changed in those four months, and thus be incorrect when applied to the former (MOE) file. A method used for some of the MCSS eligibility extract files to compensate for this is to overlay the zip code of the month of eligibility file over that obtained for the latter file if a zip code exists for a particular beneficiary and the same county in both files. (An analysis of the results of this process suggests that the CalWorks population is the most mobile, and the SSI population the least.) When this adjustment is made to the zip codes, the resultant zip code may still be incorrect even though the county is correct. This may be due to persons moving from one zip code to another within the same county or persons becoming eligible retroactively, then moving to a different county and/or zip code. In any case, beneficiaries may still have a (mailing) zip code on file that's legitimately outside the county of eligibility. These cases include: 1) Persons who have moved but haven't had their case transferred from the prior county; and, 2) Foster children or others in the custody of persons living outside the county of eligibility.

It should also be noted that some eligibility records do not have any information in the address field, a situation caused by either of the following conditions: 1) A county establishes eligibility for a person based on a one month "immediate need" criteria; or, 2) a county reports an incomplete address (e.g., no zip code), which results in MEDS suppressing all address information.

In computing ages of beneficiaries using eligibility files by subtracting the month of eligibility from each person's birth date, the fifteenth of the eligibility month should be used to provide an "average" date of eligibility.

3.3 Method for Determining Eligibility from the MMEF

Researchers outside DHS normally cannot access the MMEF. However, it may be of interest to understand how eligible counts are established. The method is as follows:

1. A person is eligible if the eligibility status code value is between '001' and '499', inclusive. (Note: See Attachment A for more details on eligibility status code values for various types of aid codes.)
2. The hierarchy of eligibility is: primary aid code, first special aid code, second special aid code, and third special aid code. Thus, if a person is eligible under both his/her primary and first special aid code, the primary is used. A SOC aid code is always in the primary position.

3. If the first, second, or third special aid category is one of the following values, it is not used: ' ', '00', '42', '46', '4C'.

4. If an aid code is one of the following, regardless of it being a primary or a special aid code, it is not used as a Medi-Cal aid code: ' ', 'IE', 'RR', '00', '50', '5G', '7X', '84', '85', '88', '89', '8A', '8C', '8D', '8F', '8X', '9A', '9G', '9H', '9J', '9K', '9M', '9N', '9R'..

As with eligibility, active status for a designated health care plan (HCP) for a given month must also be established. The HCP status code has two bytes, but active status is indicated whenever the second byte is a '1'. The values in the "Medicare Part A" field that indicate "active" status are '1,' '2,' '3,', and '7'; the respective values for "Medicare Part B" field are '1,' '2,' '4,', and '7'.

3.4 Tracking Beneficiaries Across Claims

Since 1990, the eligibility system has used the Social Security Number (SSN), or a State-assigned pseudo, as the permanent identifier of Medi-Cal beneficiaries, or MEDS ID. Tracking beneficiaries through claims using the thirteen- or fourteen-digit beneficiary identification number (BID) could result in the miscounting of beneficiaries, for several reasons:

1. Each county attempts to issue a unique ten digit number, which becomes the last ten digits of the 14-digit BID. Sometimes, however, family units (FBUs) will split, thereby necessitating issuance of a number similar to the previous one for a beneficiary, but different in the FBU and person number.

2. For other reasons not fully understood, a county may issue a completely new number.

3. A beneficiary may move to a different county, thereby being issued a completely new 14-digit BID.

4. Since the 14-digit BID is constructed differently for SSI/SSP recipients than for others (the 10-digit number of SSI/SSP recipients is comprised of a '9' and the SSN), when a beneficiary is re-categorized from or to an SSI/SSP aid category, the entire BID changes.

5. A claim received by the fiscal intermediary with an invalid BID but a valid Proof of Eligibility label or Automated Eligibility Verification number, or has been approved by Medicare, is not subject to denial for eligibility reasons; in some of these cases, it is believed the BID showing

on the claim record has an 'M' in the fifth of the 14-digit number. Cases one through four will not result in mis-counting of beneficiaries when the SSN is used to select claims.

In a preponderance of cases, a beneficiary has a single SSN and/or aid code within an eligibility month. Very infrequently a beneficiary may be assigned more than one SSN/aid code within the

same month. If a different SSN is assigned to a person, perhaps because their current one is a pseudo-SSN, that change will take place immediately on MEDS; this could occur for as many as 10,000 SSNs a month. Similarly, if a beneficiary has their aid code changed within the month to either permit the same or greater scope of benefits, this change will also occur immediately.

A California Index Number (CIN) is also issued to each Medi-Cal beneficiary. In some cases, a beneficiary may have one CIN and more than one SSN. CINs were created, among other reasons, to provide a linkage to records for the same individuals enrolled in predominately the public health programs which are not permitted to collect SSNs.

In cases where a beneficiary does not yet have an SSN (e.g., child), is not entitled an SSN (e.g., illegal immigrants), or is not tracked on MEDS by their SSN (e.g., minor consent), he/she would be assigned a pseudo-SSN. Pseudo-SSNs are issued sequentially, begin with an '8' or '9', and end in a 'P.' About 17% of all beneficiaries over a thirty-month period have a pseudo-SSN (P-SSN). In cases where a beneficiary eventually receives a permanent SSN, it is thus important to know how this may effect research that require tracking a beneficiary over time, e.g., continuity of eligibility, medical utilization of a cohort population. Therefore, we looked at the characteristics of beneficiaries on MEDS with a P-SSN who, over this thirty-month period, could be linked via their CIN number to a subsequently-issued SSN.

About 1.4% of all beneficiaries with a P-SSN over this period could be linked to a regular SSN. When claims for the P-SSN beneficiaries were pulled using either their P-SSN or their linked SSN for a sample six-month period, it was found that, under the linked SSN, 23% of their claims expenditures were for aid code **3N** (AFDC-MN-1931(B) Non CalWORKS), 12% for aid code **60** (Disabled-SSI/SSP-Cash), 8% for aid code **30** (CalWORKs-All Families), 6% for **47** (200% FPL Infant Citizen), 5.5% for **34** (AFDC-MN), and 5% for **3V** (Section 1931(b) Undocumented Aliens). For beneficiaries with linked MEDs IDs, expenditures under their P-SSN were 21% for aid code **3N**, 9% for **3V**, 9% for **34**, 9% for **47**, 7% for 30, and 7% for aid code **58** (OBRA Alien). When claims for P-SSN beneficiaries that could not be linked were reviewed, the expenditures were: 29% for aid code **3V**, 13% for aid code **48** (200% FPL Pregnant OBRA), 12% for aid code **5F** (OBRA Alien-Pregnant Woman), 11% for aid code **58**, and 6% for aid code **3N**.

Expenditures by age category for the linked-SSN, linked-pseudo-SSN and non-linked pseudo-IDs were also reviewed. The most significant difference is that for under-one year of age: for the linked SSN claims, the percent of total expenditures was about 48%, whereas for the unlinked pseudo-IDs the percent was 24%. Likewise, for the linked-SSN claims, 68% of the expenditures were for beneficiaries under 18 years of age, but only 35% of the expenditures for the non-linked P-SSN claims were for beneficiaries in this age category. The only other significant difference between claims for the linked-Pseudo IDs versus the unlinked group was that about 37% of the former claims were for CCS/Medi-Cal dually eligible beneficiaries, whereas only about 17% of the unlinked-ID claims were for these dually eligible beneficiaries.

3.5 Managed Care Enrollment Reporting

The Medi-Cal Managed Care Division (MMCD) reports counts of enrollees in plans in various reports. These counts will not agree exactly with those derived from the automated system, which is used by MCSS, due to retroactivity transactions for these plan enrollees. In the case of COHS model plans, the MMCD uses counts for the upcoming month of eligibility, then performs an adjustment every six-months to add in those eligibles who were retroactively added to (or, infrequently, dropped from) the plans' rolls. On the other hand, GMC and Two-Plan model plans have very few adjustments to the counts taken at the beginning of the month. What additions or deletions that are made are done based on all adjustments made in the last thirty days regardless of which month the transaction occurred. The automated system, however, shows the plan status for each month in history.

4. Fee-For-Service Claims

4.1 Validity, Completeness and Accuracy of Fee-For-Service Claims

FFS claims data can be very useful for many different research, fiscal and policy purposes. Because claims are processed through a fee-for-service adjudication system by the fiscal intermediary, the claims data for all critical fields are validated against fields on the same claim, e.g., gender and procedure code, as well as against external tables, e.g., procedure codes, provider numbers. There is an assumption made by some that fee-for-service data bases substantially represent all services rendered, and they are significantly accurate. Nonetheless, caution must be exercised. The services history for given beneficiaries can only begin to be complete for the months they are on Medi-Cal, so for those intermittently on Medi-Cal, adjustments must be made. In addition, there is a tendency toward "up-coding" on claims submitted for payment, that is, coding the procedure at a slightly higher level than the service level would justify to obtain a higher reimbursement rate. Fraudulent billing for services not provided is another source of error in the FFS claims data.

Other limitations are evinced by numerous studies. Some that have compared FFS claims data with medical charts for the same beneficiaries have shown that only about 80 to 90% of all services billable to medicaid are received and paid as claims.[1] Some have found that Medicaid pharmacy claims document 94% of pharmacy prescriptions.[2] Even with these problems, administrative claims data bases have been successfully used in constructing episodes of care.[3]

Not only can FFS data be incomplete, but, despite extensive auditing by the fiscal intermediary before payment is authorized, claims can be coded with different but acceptable values by providers. The use of diagnosis codes using the ICD-9-CM coding scheme is one example: the codes are variously interpreted, starting with physicians themselves, for application to the medical condition at hand, are sequenced as primary vs. secondary without use of a strict set of rules, and are influenced by the type of FFS reimbursement for the service. One study found that "agreement

in the principal diagnosis between hospitals' reports and the IOM (National Academy of Science, Institute of Medicine) reabstraction was only 65.2 percent." [4] According to the same source (Risk Adjustment for Measuring Health Care Outcomes), procedure coding generally has a higher degree of completeness, validity, and reliability than that for diagnosis codes. It is suggested in this source that this is due to a greater specificity that procedure code schemes provide, as well as the need for greater accuracy to receive FFS reimbursement. Other studies suggest just the opposite, that diagnosis codes have more reliable information than procedure codes. [5]

For a complete discussion of the limitations of administrative medical claims data bases, please refer to the MCSS report “Report on the Use of Medi-Cal Managed Care Encounter Data for Research Purposes,” available on the MCSS website.

4.2 Timeliness of Fee-For-Service Claims

Claims must be submitted to the Medi-Cal fee-for-service intermediary by providers within six months, although some claims may be accepted for processing up to twelve months after the service date but could be subject to a discounted payment (see the EDS Medi-Cal provider manuals for more information). Adjustment claims are an exception to billing limitation requirements. Realistically, given the financial incentive to obtain payment as quickly as possible plus the efficiency of electronic methods to do so, claims are submitted rather quickly. Once processed and toward the end of each month, claims payment records are sent to the State for their use. Researchers must be aware of the service-to-processed-month lag information to ensure enough months of claims data are used to get as complete a picture as possible of the services rendered over a period of time. Tables 1a and 1b provide data on lags of claims by claim type, using payment and number of claims as measures.

Table 1a, Cumulative Paid Claim Amounts, by Claim Type

Elapsed Months (rounded up)	Drug	Inpat/Hospital	Inpat/LTC	Medical	Outpat
0	14.36%	0.00%	0.06%	0.14%	0.17%
0.5	93.43%	8.52%	38.70%	26.24%	29.37%
1	95.80%	26.71%	55.30%	47.00%	48.01%
1.5	96.99%	41.76%	79.77%	61.17%	62.71%
2	97.68%	53.68%	85.63%	69.91%	72.41%
2.5	98.16%	62.32%	89.35%	75.80%	78.83%
3	98.51%	68.99%	91.66%	79.93%	83.57%
3.5	98.77%	74.16%	93.54%	83.13%	86.70%
4	98.98%	78.08%	94.97%	85.71%	89.13%
4.5	99.15%	81.33%	96.05%	87.85%	90.98%
5	99.29%	83.88%	96.85%	89.60%	92.41%
5.5	99.41%	86.03%	97.48%	91.17%	93.59%
6	99.52%	87.84%	97.97%	92.50%	94.62%
6.5	99.60%	89.35%	98.38%	93.70%	95.47%
7	99.66%	90.55%	98.68%	94.68%	96.14%
7.5	99.71%	91.62%	98.93%	95.51%	96.62%
8	99.75%	92.53%	99.10%	96.24%	97.01%
8.5	99.79%	93.42%	99.27%	96.93%	97.38%
9	99.82%	94.12%	99.40%	97.45%	97.76%
9.5	99.84%	94.73%	99.50%	97.87%	98.28%
10	99.86%	95.28%	99.58%	98.22%	98.73%
10.5	99.88%	95.79%	99.64%	98.52%	98.98%
11	99.90%	96.20%	99.69%	98.77%	99.18%
11.5	99.91%	96.58%	99.74%	99.00%	99.36%
12	99.94%	96.91%	99.78%	99.22%	99.54%

Table 1b, Cumulative Number of Claims, by Claim Type

Elapsed Months (rounded up)	Drug	Inpat/Hospital	Inpat/LTC	Medical	Outpat
0	0.00%	0.00%	0.64%	0.11%	0.28%
0.5	12.61%	8.52%	61.79%	34.01%	31.22%
1	36.17%	26.71%	73.60%	57.46%	48.36%
1.5	59.58%	41.76%	84.65%	70.31%	59.33%
2	69.59%	53.68%	89.00%	77.42%	66.79%
2.5	75.73%	62.32%	91.90%	81.97%	72.95%
3	80.11%	68.99%	93.67%	85.24%	78.29%
3.5	83.38%	74.16%	94.97%	87.70%	82.49%
4	85.88%	78.08%	95.94%	89.64%	86.00%
4.5	87.83%	81.33%	96.69%	91.23%	88.65%
5	89.42%	83.88%	97.30%	92.48%	90.59%
5.5	90.75%	86.03%	97.78%	93.55%	92.21%
6	91.83%	87.84%	98.16%	94.49%	93.62%
6.5	92.76%	89.35%	98.48%	95.29%	94.78%
7	93.53%	90.55%	98.72%	95.93%	95.64%
7.5	94.21%	91.62%	98.93%	96.45%	96.26%
8	94.81%	92.53%	99.10%	96.94%	96.75%
8.5	95.36%	93.42%	99.26%	97.40%	97.20%
9	95.81%	94.12%	99.38%	97.81%	97.63%
9.5	96.22%	94.73%	99.48%	98.16%	98.13%
10	96.60%	95.28%	99.56%	98.48%	98.57%
10.5	96.93%	95.79%	99.63%	98.77%	98.86%
11	97.26%	96.20%	99.69%	99.03%	99.11%
11.5	97.56%	96.58%	99.74%	99.26%	99.34%
12	97.82%	96.91%	99.79%	99.47%	99.57%

4.3 Physician Reimbursements

Physicians recover their costs for treating Medi-Cal beneficiaries under fee-for-service in one of two ways. First, they can bill Medi-Cal directly as the billing and rendering physician; this would usually occur for billings on a medical claim form, and would include services for a beneficiary in both an outpatient and inpatient setting. In other cases, the physician renders services while on the staff of or under subcontract with a clinic or hospital, outpatient or inpatient. While still common in the cases of outpatient facilities (e.g., clinics), it is infrequent for a physician to get his/her reimbursement directly from the inpatient hospital.

Physicians who render Medi-Cal services are not always identifiable from the claim record. If not identified from the billing provider, they can sometimes be linked through the rendering provider field. Using a sample extract of Medi-Cal claims, a valid physician number was found in either the billing or rendering field by provider type in the following percentages for the procedure code 9921X; this is shown in the following Table 2:

Table 2

Provider Type	Yes	No
Certd Family Nurse Prac	38%	62%
Community Hosp-Outpat	25%	75%
County Hosp-Outpatient	23%	77%
Nurse Midwife	90%	10%
Organized Outpat Clinic	29%	71%
Physician	100%	0%
Physician Group	100%	0%
Podiatrist	49%	51%

4.4 Adjudication Process for Fee-For-Service Claims

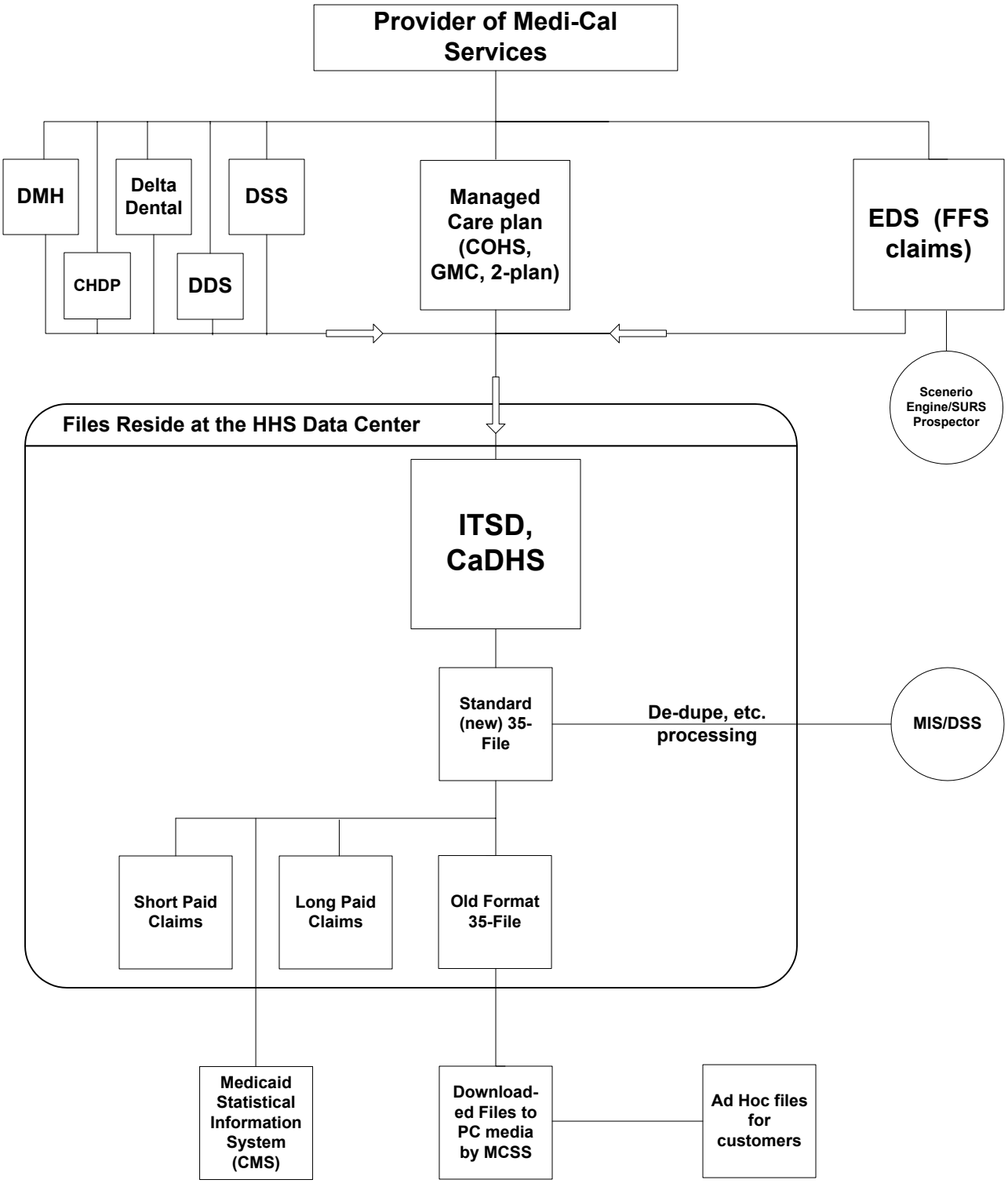
Fee-for-services claims are adjudicated, or checked for appropriateness of payment or denial, on a flow basis as they are received from providers. About 86% of all claims are received electronically by EDS, under the Computer Media Claims (CMC) program.

Except for inpatient claims, all claims are split into individual claim lines, with the common header data replicated for each line, and adjudicated in this fashion. Crossover claims are processed this way, but reconstituted into a single claim just before final adjudication; a header claim line is created for each crossover claim, under which the claim is paid (see Appendix C for more information on this subject). The paid claims tapes received by the State have all claim lines adjudicated that month artificially joined back together under its common header. (Again, hospital and crossover claims appear on the paid claims tape whole.) Claim lines for the same claim adjudicated different months will thus have the same ICN on our claims files (spanning more than one month). When using payment files unadjusted for month of service, i.e., a researcher is using just one month's

payment file, it is important to note that a month can represent four or five weeks of processed claims. Yet an "average" month will have 52 divided by 12, or 4.33, weeks of claims data. Annualizing one month's payment data should thus compensate for this (as well as such salient factors as seasonality of medical conditions). Another factor depressing the number of claims paid for a given month is the occurrence of holidays: fewer visits, surgeries, etc. take place on holidays (as with weekends) as during a normal weekday.

As Table 3, Medi-Cal Services Data Flow Diagram, below, indicates, there are many different processors of claims.

Table 3, Medi-Cal Services Data Flow Diagram



It should be noted that individuals may be entitled to receive services under both Medi-Cal and non-Medi-Cal programs. For instance, most beneficiaries in the CHDP program are Medi-Cal eligible. Thus, to obtain a complete history of medical services for dually eligible beneficiaries, the capture of claims information from all programs they receive services under is required. With regard to the CHDP/Medi-Cal program overlap, it is important to identify the overlap by beneficiary to quantify the provision of preventive care provided to the Medi-Cal population. (CHDP claim forms, unpriced, are received from managed care plans.) Those beneficiaries who are dually eligible for CCS and Medi-Cal will have all their claims paid for by Medi-Cal, whether the service relates to the CCS condition or not. Claims for some services or drugs for these beneficiaries may not be sent to the CCS Field Office for approval, and thus will not carry an indicator in the claim record as “CCS.”

(Note that many references to CCS/GHPP are by the term "CCS" only. To be precise, a beneficiary is in the CCS program if less than 21, and in GHPP if over 20. However, only about 2% of the expenditures are for GHPP clients, and 98% for CCS. For a further explanation of the CCS and other non-Medi-Cal medical programs, please refer to Appendix B.)

In the same way, persons can be a member of a Medi-Cal managed care plan, but can still receive their carved-out services under FFS. Such FFS claims can be identified by the health care plan number appearing in the respective field of the FFS record.

Not only can beneficiaries be eligible under more than one program, they can also be simultaneously eligible within Medi-Cal under more than one aid code. On the MMEF, the additional aid categories are posted to the first, second and third aid code positions. The EDS claims processing system checks for eligibility under all aid code positions (i.e., primary, first through third special) and picks the best one to use based on both full versus limited scope characteristics and the amount of FFP attached to the aid code to maximize FFP.

4.5 Adjustments and Accounts Receivables Applied to Fee-For-Service Claims

Since providers have between six and twelve months to submit claims to the fiscal intermediary from the date services are rendered, claims for a given date of service may not be received for several weeks. Once received, the claims could also be returned to the provider for missing or revised information, adding more time to the adjudication process.

Claims (actually, claim lines, excepting inpatient and crossovers) are sometimes voided or adjusted after being paid. When a void occurs, a record identical to the first (paid) claim is created, but with negative reimbursement and units amounts; thus, when "added" to the first claim, the resulting amount is zero. An adjustment starts with a void, then a third line is created with the correct reimbursement and units values. The latter two claims will have the same claim number (ICN), with different adjustment codes.

Claims are also sometimes "paid" at zero reimbursement amounts; this may occur in the case of Medicare crossovers when the residual owed by Medi-Cal is zero or negative, and in the case of a provider who incurs a Medi-Cal accounts receivable due to perhaps an overpayment. *Caution:*

When tracking reimbursements costs, remember to recognize that zero paid claims often occur for beneficiaries over sixty-four years of age.

The rate of voids for pharmacy claims is relatively high due to two reasons: 1) Beneficiaries might not pick up their prescriptions after a pharmacist has prepared them in advance and submitted them electronically for payment, and thus the pharmacist must then return the drug to stock and request of EDS a reversal of the original claim; and, 2) Pharmacist providers can bill in real time, so if a mistake is made, the original claim that has already been electronically submitted must be reversed to avoid the second (correct) one from being denied as a duplicate by the fiscal intermediary.

Claims data do not include accounts receivable, cost settlement, or advanced payment offsets, which are not applied to individual claims within the claims processing system. Claims are processed normally with the paid amount based on whatever EDS's system shows as the allowable amount, less copays, etc. Then, on the payment tape sent to the State Controllers Office, any offsets are applied against the total check that the provider would have gotten for all the claims processed that week.

4.6 Fee-For-Service Claims Files

MCSS uses primarily the original version of the 35-file for the source of its claims data, as opposed to the newer Standard 35-file, the Short Paid Claims and Long Paid Claims files. MCSS also uses such support files as the procedure code file, the drug file and the provider master file. Eligibility files are also extensively accessed, often with the respective claims files.

Of all the claims files, the one with the most information is the 35-file. The kinds of valuable information this file has contained since March 1994 is: 5-byte primary and secondary diagnosis codes, primary and secondary surgical codes for inpatient claims, prescribing/referring provider numbers, plan codes for beneficiaries in managed care plans on the date of service, and NDC (drug) codes.

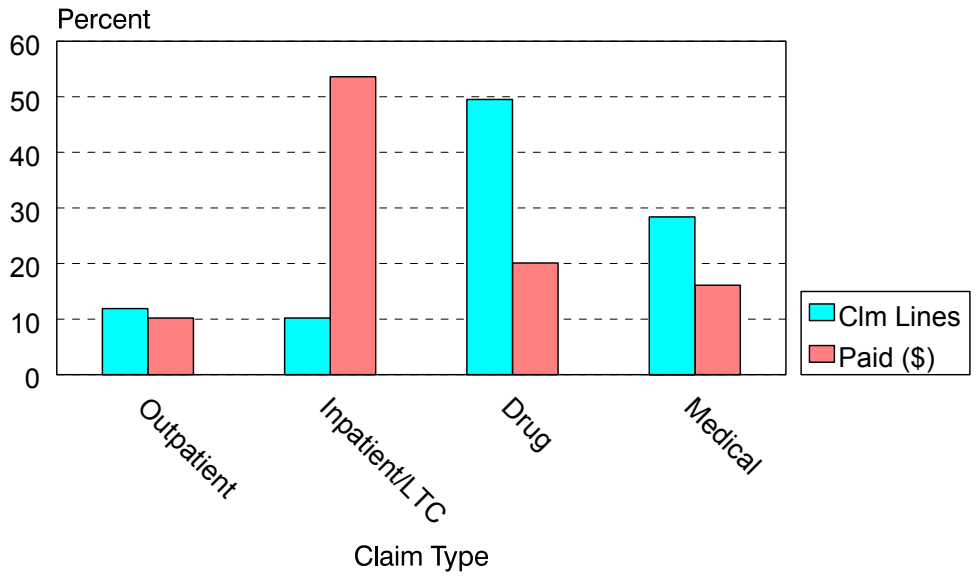
Special note should be made about the drug code DHS uses, the NDC (National Drug Code). The NDC description, Smartkey, etc. is available to DHS through a contract our fiscal intermediary, EDS, has with the First Data Bank company. Use of this information is limited to use for the administration of the Medi-Cal program. Thus, only the NDC code (which is received on the claim from the providers) can be provided to organizations outside DHS. However, most of the descriptive information (excepting the Smartkey) is available from various sources via the Internet.

4.7 Fee-For-Service Claim Type Proportions

Claim types vary in the characteristics of their volumes and the proportion of the Medi-Cal dollar they consume. Chart 3, Claim Lines vs. Dollars by Claim Type, below (produced in January 1999)

indicates how volumes of claim lines and reimbursed Medi-Cal dollars can vary depending on the type of claim it is.

Chart 3, Claim Lines Vs. Dollars by Claim Type



4.8 Utilization Intensity by Calendar Month

Services are not used at the same rate throughout the year by Medi-Cal beneficiaries. Using a 10% sample of claims, we looked at the intensity of service utilization in Medi-Cal for the calendar year 2003 by major program.

The first method used was to calculate an average number of users per month based on the annual number, then convert each month's number of users to a percent of the average monthly. Calculating a standard deviation by program provided the following:

Table 4

Plan no. and Program	Std. Dev.
6-DSS	0.010
1-DDS(Waiver)/DSS(PCS)	0.017
9-EDS(FFS)	0.043
8-DMH(Short-Doyle)	0.055
0-Dental	0.072
5-EPSDT	0.164

To determine which months had the highest or least utilization, two methods were used, each being applied only to the top twenty vendor codes (by utilization): 1) assigning weights by vendor code for each month that had the highest rank number of users, the second highest, etc., then using these rankings to derive a score for that month based on the highest number of utilizers; and, 2) assigning scores based on how many rank 1, 2 and 3 scores each month had for the twenty vendor codes. The resultant scores indicate that the highest utilization months are (ranked hjgh to low) January, October and March. The lowest utilization months are August, June and, lowest of all, November. According to one of the methods, September is the fourth highest utilization month, and according to the other method, December is. (December is the highest utilization month for drugs.)

It's interesting that the top ranked months, January, October, and March, each have 31 days during the month, and there is only one holiday for the three months combined, New Year's Day (King's birthday is not universally celebrated). Three of the four lowest ranked months are August, June and November, months that are either partially or wholly school vacation months or summer months, or which have some holidays (November has Veterans' Day and Thanksgiving, for which two holidays are usually taken). In addition, June and August have only 30 days in the month.

When the standard deviation for utilization is derived by claim type, we get the following:

Table 5

Clm Type	Std. Dev.
2-Inpat	0.017
3-Drug	0.027
4-Medl	0.054
1-Outpat	0.060
5-Dental	0.072

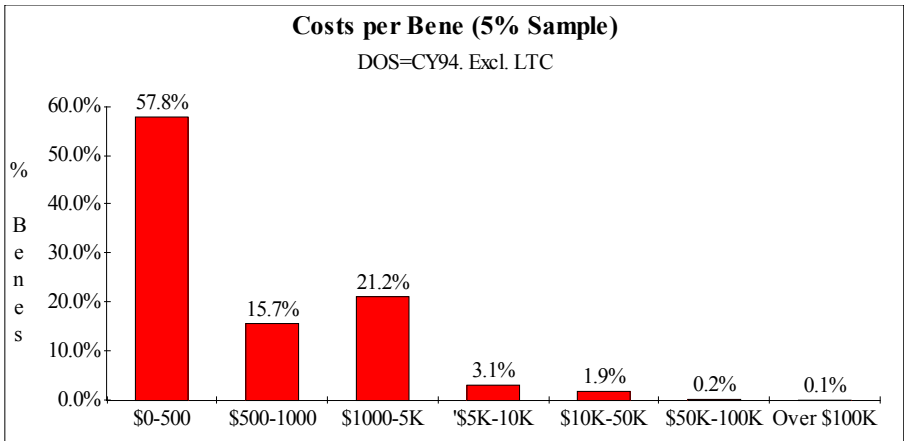
This ranking suggests that there are some medical conditions for which a delay in treatment can be tolerated by the beneficiary, and this delay might more likely occur in months with more holidays.

Other factors, such as summer vacations, flu seasons, number of days in the month, etc. also seem salient.

4.9 High Cost Beneficiaries using Fee-For-Service Claims Data

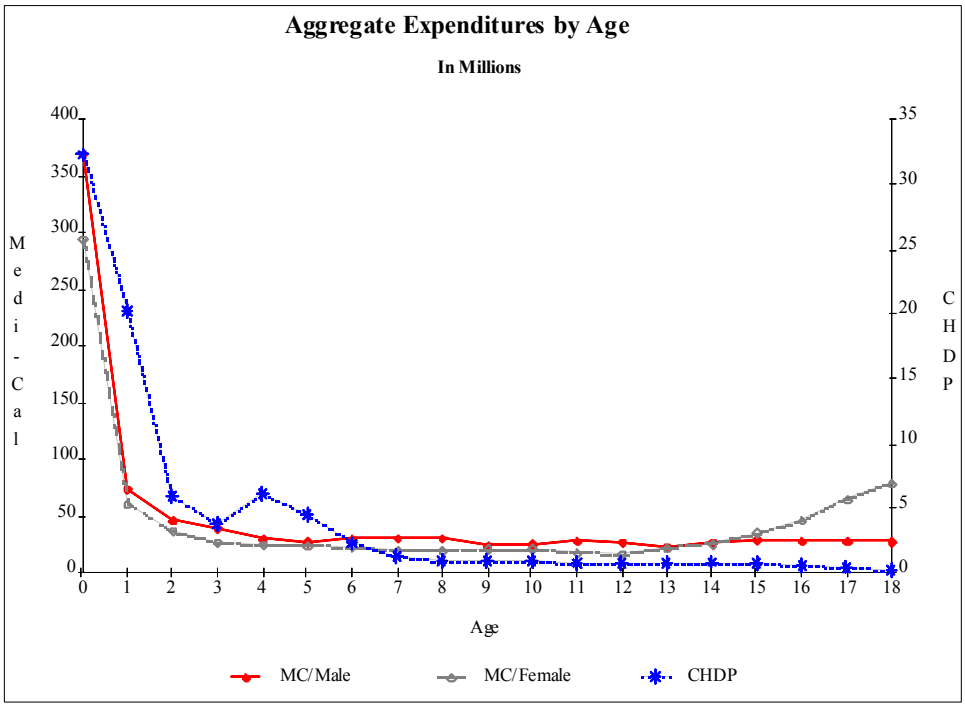
When analyzing fee-for-service claims data, it is sometimes important to compensate for healthy beneficiaries. In such cases, using eligibility files to obtain the "denominator" of all beneficiaries eligible for benefits under FFS during a given month is critical. On the other end of the spectrum are beneficiaries who use a disproportionate amount of Medi-Cal resources. Preliminary data suggest that as few as 3% of the non-LTC beneficiaries use about 50% of the Medi-Cal dollars over a given six-month period, and 10% use about 75%, as Table 6, Costs per Beneficiary (5% Sample), illustrates.

Table 6



In general, beneficiaries consume fewer Medi-Cal resources in terms of aggregate dollars and dollars per capita the older they are. This probably relates to the fact that the disabled are usually older but require many services due to their condition. The notable exception is the group less than about three years of age. Aggregate Medi-Cal expenditures for those under eighteen years of age, by year, are shown in Chart 4, Aggregate Expenditures by Age.

Chart 4



As this chart shows, for both the Medi-Cal and CHDP programs, costs for those under one year old are relatively very high, even though the CHDP program does not cover the traditionally expensive hospital care. Also notable is the much higher costs for males under Medi-Cal (about 50% more at ages three, seven, and eight) than for females; however, by age fifteen, female costs exceed aggregate male costs (due to delivery costs).

Researchers can use Medi-Cal utilization reports to ascertain the levels of utilization for specific medical services, in units, days, or dollars per thousand. Such data must be treated with caution, for these reasons: 1) Persons sometimes become eligible for Medi-Cal to have medical conditions treated, then become ineligible after becoming healthy, thereby making their utilization rate while on Medi-Cal appear high; and, 2) Medicare beneficiaries for which Medi-Cal pays the residual will be counted as beneficiaries as will their services even though Medi-Cal pays a very small percent of their overall care.

In general, there are about 30 claims for each eligible beneficiary over a given year; that is, if a county had 100,000 beneficiaries, the number of Medi-Cal claims to be expected to be paid for that county over a year's period would be about 3,000,000.

Viewed a different way, about two-thirds of the beneficiaries are in aid categories included under expanded managed care (generally, the CalWorks family groups); however, this group incurs but one-third of the Medi-Cal dollar.

4.10 Three Methodological Algorithms Using Fee-For-Service Claims

Here are three methodological algorithms MCSS commonly employs using fee-for-service claims data.

1. **Inpatient Length of Stay** -- One of the more frequent uses of hospital claims data is determining length of stay. The criteria for deriving this information is:

    a. Select claims that are inpatient (claim type = 2) hospital (vendor code = 50 or 60), non-adjusted (adjustment indicator = 0).
    b. Add total number of days from the header.
    c. Add total number of discharges (claims with discharge not equal to 6).
    d. Divide the number of days by the number of discharges.

2. **Identifying Newborns from Claims Data** -- Services for newborns, particular for inpatient claims, can be billed on the mother's ID for up to two months after birth. A newborn is, then, identifiable in the claims data by checking for a birth date that is either the same month of the delivery, or the month following the date of service. (See Appendix C for a more thorough discussion of birth dates on claims.) Establishing denominator beneficiary counts is somewhat more difficult, however. A quick estimate could be made by performing a non-linear regression for the beneficiaries with ages four months and less based on counts of beneficiaries from five to twelve or so months. (Note: Total beneficiary counts would thus increase in the amount of the difference between actual counts and counts found from these projections).

3. **Identifying an Eligibility Denominator for Use with Claims Data** – Often times researchers want to know how much a beneficiary costs when on Medi-Cal. Several factors should be recognized when determining a cost per month per beneficiary.

    a. Exclude from the claims costs those beneficiaries in aid codes for whom we have no or little eligibility data. These particularly include Family PACT (8H) and presumptive eligibility (7F and 7G).

b. When costing eligible beneficiaries by age, if possible, use the birth date from the eligibility file, a field edited by MEDS, rather than the claim, which sometimes has errors (I found about 5% of the claim birth dates result in invalid ages, and the 65+ age category is overstated significantly as a percent of the whole using claim birth dates); the exception is the case of newborns using the mothers ID, then use the newborn's claim birth date. Note that newborns with the mom's ID are not on the eligibility file, so caution must be used in developing a cost per eligible per month for those under 12 months in age.

c. Exclude from the claim costs and the eligibility files those in managed care, unless these costs are combined with the managed care capitation payments. Some of the costs for the managed care enrollees are FFS when they are carve-out, non-capitated services.

4.11 Using Birth Dates

Analyses of claims often require beneficiary age as a variable which, in turn, must be calculated using the beneficiary birth date and date of service. Since the birth date on the claim is not highly edited, unlike the birth date on the Medi-Cal eligibility files, it is important to know when to trust the birth date from the claim.

A birth date on the claim may be correct even though it does not match that on the eligibility file. On the other hand, the birth date on the claim can probably be trusted even though there is no matching record, by SSN and month, on the eligibility file.

Considering first those situations where a matching eligibility record cannot be found, for a recent three month period (first quarter of 2005), 89.5% of the birth dates on the claims matched exactly to that on the eligibility file, 2.3% did not match exactly, and fully 8.2% of the claims did not have a matching eligibility record. Looking at the 8.2% of the records that could not be matched at all, 7.77% of this percent are attributable to the beneficiaries that, by the nature of the program under which they received services, did not have an eligibility record; these programs are the Family PACT and the Presumptive Eligible programs. If the claims for these programs are removed from our analysis, 97.1% of the claim birth dates match exactly to the eligibility file, 2.4% match in some way, and only 0.5% of the claims not having any counterpart eligibility record.

The percent of matching between the claim and eligibility record by category is shown in Table 7.

Table 7

Claim/Eligibility Record Category	Percent of all Claim
Exact match betw eligible and claims	97.07%
Both dates less than 12 mos	0.01%
Match on 0 digits	0.01%
Match on 1 digits	0.04%
Match on 2 digits	0.09%
Match on 3 digits	0.09%
Match on 4 digits	0.08%
Match on 5 digits	0.21%
Match on 6 digits	0.89%
Newborn using the claim birth date	0.88%
Newborn, using the ICD9 code	0.04%
No Claim Birthdate	0.08%
No eligible record by MEDS ID and month	0.50%

For the last category, "No eligible record by MEDS ID and month," 51% of the 0.5% were for pharmacists, 14% were for physician groups, and 12% were for Rural Health/FQHC/Indian Health providers. By aid code, the beneficiaries with the no-eligibility-record claim were 3N (21%), 8W (17%) and 8U (13%).

5. Managed Care Encounter Claims

Managed care encounter data are different from fee-for-service Medi-Cal claims in many respects. For instance, fee-for-service (FFS) claims must be submitted to the fiscal intermediary within one year of data of service. The County Organized Health System plans, on the other hand, do not have any such requirement in terms of submittal of their claims to the State. The Two-Plan and Geographic Managed Care (GMC) plans must submit their encounter data to the State within ninety days of the date of service, although the State is not monitoring this requirement. The requirements for coding is virtually the same between the FFS and COHS claims. However, the Two-Plan and GMC plans have somewhat different coding requirements, and thus the researcher must carefully review these before using their encounter data. For instance, provider numbers may be that used for Medi-Cal FFS, the license ID, a facilities ID, or a tax ID. Modifier values may include several unique to Two-plan and GMC plans.

5.1 Validity, Completeness, Accuracy and Consistency of Encounter Claims

As with FFS claims, managed care encounter data can be useful to policy-makers and researchers alike for a variety of uses, including assessing quality of care, effectiveness of health protocols and technologies, and assessing the effectiveness and efficiency of the managed care plans. However, managed care encounter data differ from fee-for-service claims data in that there is no direct financial incentive for providers or managed care plans to submit complete and accurate documentation. This fact requires researchers to be aware of how incomplete and inaccurate managed care encounter data are before using them. These limitations are in addition to many of those documented for fee-for-service data, such as imprecise coding for diagnosis and procedure codes and missing claims for periods of ineligibility for given beneficiaries. Encounter data for the incurred period CY99 were recently reviewed on the characteristics of validity, completeness, accuracy and consistency. A summary of the results is as follows.

Validity. Validity tests conducted on CY99 managed care encounter data resulted in very low error rates for diagnosis codes (0.8%), procedure codes (0.8%), procedure modifiers (1.5%), national drug codes (0.69%), provider types (virtually error-free), and physician specialty (0.6%). Some plans had much higher rates of invalid values, but usually less than 10%. Because the list of permissible valid modifiers is substantially different between the COHS and the Two-Plan/GMC plans, researchers must be careful to adjust for this. The frequency in the use by some plans for values equating to “miscellaneous/other” for provider type and specialty is abnormally high (4 to 30%).

Most encounter data records have provider numbers that cannot be validated, either because they are not listed in files maintained by the Department of Health Services or because the Department does not have access to a master list. Only 16.0% of the records had a provider number that could be matched to the Medi-Cal provider master file (28% non-drug, 3.5% drug), and 6.5% had a provider number in the form of a license number that was matched to this provider master file. An additional 4.8% of the provider numbers were matched to lists provided by the managed care plans or to the license number file provided by Consumer Affairs. Other forms of the provider identifiers permitted for use by Two-Plan/GMC plans, such as facility, tax and national identification numbers, could not be checked for validity. When the information on provider name, provider type, and zip code given on the encounter record was matched to the same information on the provider master file, the successful match rates varied from 34% to 92%. Information for these three encounter data variables (provider name, provider type, zip code) was evaluated to determine if they were unique to the provider number given by each of the plans in their encounter data. On average, for each unique provider number given by a plan, there were 1.8 unique names, 1.1 unique provider types, and 1.3 unique zip codes. The inability to identify most providers rendering services to managed care recipients seriously undermines the use of encounter data for such purposes as monitoring access to providers with certain specialties or cultural/linguistic characteristics, are in certain geographic locations, or are traditional, safety-net providers.

Beneficiary identification numbers in the form of the MEDS-ID (usually social security number) were verified against eligibility records for the correct month and plan assignment 97.9% of the time for all plans; eliminating one plan with a high error rate increased the rate for the other plans to 98.7%. Such a high rate of valid beneficiary identifications would permit the encounter database to be used for longitudinal analyses and creation of samples files for research purposes.

Accuracy. Encounters for drugs were analyzed for accuracy by comparing the distribution of drug encounters for managed care to FFS using a drug classification scheme known as general therapeutic class. There was significant variation in the percent each class comprised of the total between the two FFS databases, FY94-95 and CY99. The percentages by class for the managed care data usually fell within the range set by the two percentages for the two FFS base years. Exceptions may be due to inaccurate data, but more likely reflect differences in practice patterns for managed care in these counties. The accuracy of diagnosis coding for asthma was reviewed by comparing the percent of encounter claims with these codes to that for FFS. Asthma was selected because it is an illness which is not easily cured, but which must be treated on a continuous basis. Although there was some variation by plan, the percent of the records in the managed care versus the FFS94-95 databases was very similar, at just over 2%. This may suggest that encounter records are usually accurately coded for diagnosis.

Completeness. Encounter data completeness was reviewed by comparing the rate of encounters per thousand beneficiaries to FFS data for FY94-95 and CY99, with counts expressed as a rate of encounters per thousand beneficiaries. Because some suggest that managed care plans successfully report at least one encounter per visit but not necessarily the second or third due to the paperwork burden, also a comparison was performed using the number of encounter days per beneficiary. The results were substantially the same whether the FY94-95 FFS or CY99 FFS data base was used, or encounters or encounter days were used as a unit of measure. Encounter data had about 33% of the FFS encounters for inpatient, 38% of the medical/outpatient encounters, and 76% of the drug encounters. Because the drug completion rate for the managed care encounter data is about twice what it is for inpatient and medical/outpatient, and because some of the managed care plans submitted many more drug claims than under FFS, persons from the managed care plans and drug claim processing organizations were consulted on this. The likely explanation for the higher rate of drug encounters under managed care is that the days supply reflects the industry norm of thirty days rather than about sixty days under Medi-Cal FFS. The observance of this norm for Medi-Cal managed care is reinforced by the caution exercised by the managed care plans not to provide drugs for future periods in which their member may no longer be on Medi-Cal or in their plan. Researchers looking at the role drugs play in managed care versus FFS should be cognizant of this difference.

Consistency. The encounter data were reviewed for how consistent they are submitted by plan on a month-by-month basis by looking at the three lowest encounter volume months versus the three highest volume months. Recognizing that demand for medical services is higher some months due to seasonal variation, e.g., allergy and flu seasons, the overall utilization should be approximately the same month to month. For encounter data, the lowest three months by plan only had about

50% of the encounters as the three highest months, suggesting that there are gaps in the data submitted by the plans. (All data were analyzed on a date of service basis.)

5.2 Timeliness of Managed Care Encounter Claims

Encounter data from the Two-Plan and GMC managed care plans must be submitted to the State within ninety days of the data of services. The County Organized Health Systems plans have no timeliness contractual requirements. A review was recently done of the CY99-incurred encounter data and fee-for-service claims data to determine the approximate number of days from the date of service of encounter claims to the date the record gets into the Department's claims files. Probably due to a high volume of encounter data received from the Two-Plan/GMC plans in late CY99, the date-of-service to processed-month for drug encounters was very high for the first two quarters of CY99. Lag periods for the rest of the data were more consistent, although they generally improved the second half of the year. Assuming future encounter data will have lag periods similar to that found the second half of CY99, the average for these two quarters (covering July through December 1999) was used to estimate a date-of-service to processed-month lag periods for encounter data, and these were then compared to lag periods for fee-for-service data covering all of CY99. These lag periods are shown in Table 8.

Table 8

	Managed Care			FFS
Claim Type	All	COHS	Non-COHS	
Drug	100	87	102	21
Inpatient	112	103	113	58
Medical	114	122	113	64
Outpatient	102	104	101	84

Given these lag periods, and considering the fee-for-service lag curves presented in Section 4.2, Timeliness of Fee-For-Service Claims, it would appear that a researcher should wait at least a year to obtain 80 to 90% date-of-service encounter data for a given month.

5.3 Submission and Processing of Medi-Cal Encounter Data

Managed care encounter data are submitted electronically in different formats and to different entities, depending on the type of managed care organization (MCO) submitting them. County Organized Health System (COHS) plans submit their encounter data electronically directly to the Department in a file format known as the "35-File." This file format is mainframe-compatible, variable-blocked, in an EBCDIC character set, and includes data formats unique to a mainframe environment, such as binary and packed-decimal. Since 1988, when the San Mateo and Santa Barbara COHS plans started submitting data to the State, the task of reviewing the files has been that of the Medical Care Statistics Section (MCSS) and the Information Technology Services

Division (ITSD). These reviews have largely consisted of detecting major problems, especially those that might cause "fatal" processing errors. A detailed and structured review of COHS data for completeness and quality has never been undertaken. COHS plans are not capitated for providing CHDP services, rather these services are rendered on a fee-for-service basis and billed directly to the Department as for beneficiaries in non-managed care counties.

Two-Plan and Geographic Managed Care (GMC) plans submit data directly to the Medi-Cal fiscal intermediary, Electronic Data Systems (EDS) using electronic media, e.g., cartridges, tapes, or online. The files are edited by EDS for three types of errors: "critical error," "one-percent error," and "five-percent error." "Critical errors" which cause a file to be rejected include: no header record present; invalid submitter ID code; record count on transmittal does not match record count from file; any single record has incorrect plan code, claim type (e.g., medical), too many segments per inpatient claim, and invalid adjudication status code (e.g., paid). "One- and five-percent error" editing involves determining the percent of validity errors for specified data elements and then, if the one- or five-percent error threshold is reached on any of these data elements, the whole file is rejected. For instance, provider type is a "one-percent error" data element. Thus, if a file had more than 1% of its provider type values that were invalid (i.e., not in the table of acceptable provider type values), the whole file is rejected. "One-percent error" threshold data elements include: those that must have values specified in the data element manual (provider type, long term care accommodation and ancillary codes); must be numeric (paid amount, Medicare deductible and coinsurance amounts, and days stay); or must meet other relational edits (beginning date of service must be earlier than ending date of service, and vice-versa). "Five-percent errors" include numeric checks (billed amount and procedure quantity), and checking of values contained in external Medi-Cal files (beneficiary Medi-Cal number, birth date, National Drug Code), and other checks.

The Two-Plan and GMC managed care plans are capitated for providing CHDP services, and thus either the plans or their providers submit Information-only CHDP claims (PM160) to the fiscal intermediary, which captures the claims information electronically and sends a tape with these data to the Department. Only superficial editing is done on these claims.

Two-Plan and GMC encounter data files which pass these checks are sent each month to the Department's Information Technology Services Division (ITSD), which reformats them for the 35-File. These records are combined with those from the COHS and are sent to the Management Information System/Decision Support System maintained by The MEDSTAT Group. About May 2001, ITSD will begin applying duplicate record logic, then delete those records which fail this edit. This duplicate checking logic includes, among other fields, the claim record number assigned by the managed care plan. The review of the CY99 encounter data included an elimination of duplicate records; this logic was almost identical to that used for fee-for-service claims: a claim was considered a duplicate if it was not an adjustment, and had the same values for the fields 1) provider identification number, 2) beneficiary identification number, 3) date of service, 4) procedure/modifier or National Drug Code, and 5) units. The rate of duplicates ranged from 9% for the first calendar quarter, to 3.6% for the last quarter. Researchers are therefore urged to be aware that Medi-Cal encounter data contain some duplicates which should be deleted before further processing.

5.4 Adjustments for Managed Care Encounter Data

As with fee-for-service claims, encounter data may also be adjusted through a void, or a void plus an adjusted claim to increase the units/paid amount, or to decrease these fields. (See the discussion of adjustments in Section 4.4.) Claims (actually, claim lines, excepting inpatient and crossovers) are sometimes voided or adjusted after being paid. At this point, claim adjustments were deleted from the databases. In the review of encounter data, out of the twenty-one processed months used for CY99 incurred data, December 1999, January 1999, April 2000, and July 2000 had a significant percent of adjustments, 12%, 45%, 37%, and 8%, respectively. The other months seventeen had an average of 1.4% adjustments of all records. It thus appears that the number of adjustments will vary substantially, perhaps until more routinized processing of encounter data can be achieved by the plans.

5.5 Claim Type Proportions

Claim type proportions for managed care encounter data is similar to that for fee-for-service (see Section 4.7, above), with some important exceptions. Because managed care is implemented in the more populous counties in California, any differences between city and rural areas will be reflected in the encounter data. For instance, Medi-Cal beneficiaries in rural areas obtain a higher percent of their care from physicians through outpatient facilities rather than physicians' offices. In addition, in the assessment of managed care encounter data for CY99, it was found that managed care plans have about twice as many prescriptions per beneficiary as fee-for-service. The best explanation for this is that managed care follows the prevailing practice of providing a thirty-day supply of drugs per prescription, whereas the practice in Medi-Cal fee-for-service tends to be providing a sixty-day supply.

Other differences in managed care from fee-for-service documented in several studies suggest that inpatient utilization should be less under managed care, and, controlling for days supply variables, drug utilization may be slightly higher if this can substitute for frequent office visits.

6. PROVIDERS

6.1 Medi-Cal Provider Number Assignments

Shortened versions of the Provider Master File (PMF) are frequently used in researching claims. It should be noted that the provider type shown on the PMF differs from the vendor: provider type is the current coding scheme used in adjudicating, whereas the vendor code is an older scheme discontinued in the late 1970's but still used for continuity purposes and for providers on claims not adjudicated by EDS and not necessarily on the PMF, e.g., Department of Aging. (The physician specialty coding scheme has not changed.)

Provider numbers are assigned by the Provider Enrollment Unit, Payment Systems Division, DHS. Non-contract (non-CMAC) hospital and outpatient provider numbers are built on a common five digit

number comprising bytes five through eight of the provider number; that is, a non-contract facility can be identified across provider types by this five byte common number. Contract (CMAC) providers have a common number comprised of bytes four through nine. (When using this method to identify common facilities, it is recommended that provider county also be used.) Solo practitioner physician provider numbers are built on the license number issued by the Board of Medical Quality Assurance.

When building the provider number, the Provider Enrollment Unit, Payment Systems Division uses meaningful provider prefixes, i.e., first three or four characters of the provider number, and one-byte suffixes. (The coding scheme and some explanation for these prefixes and suffixes is available in the 35-File Manual.) Application of this coding scheme results in one facility being assigned different provider numbers, depending on what service is being rendered. Similarly, one physician may receive multiple provider numbers, one for each practice site and/or specialty; in one analysis, 15% of the active physician provider numbers had a duplicate first-eight-byte provider number (approximately the same for groups and solo physicians).

6.2 Provider Characteristics

As with beneficiaries, providers can vary tremendously in the relative amount of reimbursements they receive from Medi-Cal. Before most managed care plans were implemented from 1996 forward, estimates were derived which indicated that only about one percent of the active Medi-Cal providers render over a third of the services using Medi-Cal reimbursement amounts as a measure, and 10% account for 78% of the reimbursements. (Also see the Annual Statistical Report for Medi-Cal earnings frequency by provider category.) Similarly, the number of CalWorks and medically needy/indigent population that providers see as a percent of their total Medi-Cal business varies; however, there seem to be more Medi-Cal providers seeing none of this population than there are providers seeing exclusively this population relative to their total Medi-Cal clientele.

Under Medi-Cal, solo physicians and physician groups are classified by specialty, as self-reported by them (rather than taken from board certification files). When coding a physician/physician group's Medi-Cal application, when not declared, a specialty is coded "general practice." Because individual physicians within group practices may have a specialty different than that declared by the group, caution must be used in counting specialties for groups. Notwithstanding this, we feel specialty information in this field is substantially valid. In terms of solo physicians and groups, each has about the same proportion (16%) of generalists (i.e., general and family practice) vs. specialists.

Another characteristic of Medi-Cal providers is that the Medi-Cal population they serve do not always live in the same county as the provider. In fact, beneficiaries see providers outside their county for about 12% of the services they receive (as measured by the Medi-Cal paid amounts (Source: Medi-Cal paid claims files for January 1998, excluding COHS, dental and mental health claims). Generally, the beneficiary and provider county will differ when the service is rendered under the following circumstances: 1) in contiguous high population counties; 2) with providers in

highly specialized professions, e.g., audiologists; 3) with providers who, due to few numbers, are widely scattered, e.g., groups; 4) with providers for whom the beneficiary does not require personal interaction, e.g., labs; 5) by providers who are very mobile in their practice and thus do not have an identifiable service location, e.g., nurse midwives, medical transportation; and, 6) for providers who render elective services to beneficiaries who thus have time to travel further to their locations.

7. OTHER INFORMATION AND FILES RELATED TO MEDI-CAL

7.1 Files and Information Available through MCSS

Those using paid claims data should also refer to the Short Paid and 35-File Claims File documentation, the Annual Statistical Report, the Services and Expenditures Month-of-Payment Report, and the Medi-Cal Provider Manuals. (All but the Provider Manuals are available from MCSS; the latter is available from EDS at http://www.medi-cal.ca.gov/ .) Medi-Cal managed care information is available from the Managed Care Annual Statistical Reports, at http://www.dhs.ca.gov/MCSS.)

7.2 Files from Other Agencies

Other State offices and agencies maintain data bases with information on Medi-Cal beneficiaries. Besides the Public Health Program FFS claims (see above), the Vital Records and Statistics Section maintains annual data bases on births and deaths for the general population. One indicator on this file is Expected Source of Payment, Medi-Cal being one possibility. The Office of Statewide Health Planning and Development (OSHPD) maintains beneficiary data bases on discharges from hospitals, again for the general population with a file indicator of Medi-Cal being a source of payment.

ENDNOTES

[1] Steinwich, Donald. October 24, 1993. Handout entitled "Medicaid Claims Data," APHA Conference, San Francisco, CA.

[2] Bright, R.A., J. Avorn, and D.E. Everitt. 1989. "Medicaid Data as a Resource for Epidemiologic Studies: Strengths and Limitations." Journal of Clinical Epidemiology 42 (10): 937-45

[3] Wingert, et al.. 1995. "Constructing Episodes of Care from Encounter and Claims Data: Some Methodological Issues." Inquiry 32: 430-443.

[4] Iezzoni, Lisa I., ed. 1994. Risk Adjustment for Measuring Health Care Outcomes. Ann Arbor, MI: Health Administration Press.

[5] Kashner, T.Michael. Fall 1998. “Agreement Between Administrative Files and Written Medical Records.” Inquiry 36(9): 1324-35

APPENDIX A

Medi-Cal Description and Expenditures/Eligibles Sources

The following is an excerpt from the Medi-Cal Provider Manual, available through the website http://www.medi-cal.ca.gov/:

> "Medi-Cal, California's medical assistance Medicaid program, became effective in March 1966. In July 1965, two major amendments to the Social Security Act greatly expanded the scope of medical coverage available to various segments of the population. Title XVIII established the Medicare program, and Title XIX established the State-option medical assistance program known as Medicaid that provided Federal matching funds to states implementing a single comprehensive medical care program.
>
> California legislation implementing the Title XIX program was signed in November 1965. Under the provisions of Title 22 of the *California Code of Regulations*, the State Department of Health Services administers the Medi-Cal program and has statutory responsibility to formulate policy that conforms to Federal and State requirements.
>
> The objective of the Medi-Cal program is to provide essential medical care and services to preserve health, alleviate sickness, and mitigate handicapping conditions for individuals or families on public assistance, or whose income is not sufficient to meet their individual needs. The covered services are generally recognized as standard medical services required in the treatment or prevention of diseases, disability, infirmity or impairment. These services are comprehensive and provide care in the major disciplines of health care."

For a description of the many Medi-Cal and non-Medi-Cal fee-for-service health coverage programs, please refer to the Medi-Cal fiscal intermediary website http://www.medi-cal.ca.gov/, click on Publications, then for any of the provider manuals, e.g., Medical, search on the term "Programs Overview."

The following table provides a delineation of Medi-Cal and non-Medi-Cal programs or program components and where the service and eligibility data reside.

Table A-1, Health Expenditures Sources

Medi-Cal

Program	Agency Administering Funds	In Month Of Payment Rept (Yes/No)	Data Files*	Vendor Code	MSIS
Medical (e.g., physician, outpatient, inpatient, drug, allied, DME, lab, HHA)	DHS	Yes	EDS	--	Yes
Drugs, Psychiatric and AIDS	DHS	Yes	EDS	--	Yes
Adult Day Health Care	DHS	Yes	EDS	01	Yes
Glass Lenses from Prison Industries	DHS	Yes	EDS	11	Yes
Genetic Disease Testing	DHS	Yes	EDS	04	Yes
Long Term Care/Mental Hlth Accomdn Svs	DHS	Yes	EDS	80	Yes
Long Term Care (Not Mental Hlth)	DHS	Yes	EDS	80	Yes
EPSDT Supplemental (Non-CHDP)	DHS	Yes	EDS	82	Yes
Local Education Agencies	DHS	Yes	EDS	55	Yes
CCS/GHPP (Medi-Cal only)	DHS	Yes	EDS	--	Yes
Dental	DHS-PSD/Dental	Yes	Non-EDS	27	Caps, not clm payments
Waiver Services: Model NF Waiver**	DHS	Yes	Non-EDS	71	Yes
Waiver Services: Multipurpose Senior Services Program	Dept of Aging	Yes	EDS	81	Yes
Waiver Services: Nursing Facility Waiver**	DHS	Yes	EDS	71	Yes
Waiver Services: AIDS	DHS/AIDS	Yes	EDS	73	Yes
Waiver Services: In-Home Medical Care	DHS	Yes	EDS	71	Yes
Waiver Services: DDS/HCBS	DDS	No	Non-EDS	76	Yes
Waiver Services: NF A/B & NF Subacute waivers	DHS	Yes	EDS	71	Yes
DDS TCM - DDS	DDS	No	Non-EDS	93	Yes
Medi-Cal TCM - DHS	DHS	No	Non-EDS	92	Yes
Drug Rebate (credit)	DHS/Drug Unit	No	Drug Unit	--	No
In Home Supportive Services/Personal Care Services Program	DSS	No	DSS; Non-EDS	89	Yes
CHDP (EPSDT)	DHS (CMS)	No	CHDP; Non-EDS	94	Yes
Short-Doyle	DMH	No	Non-EDS	64/95	Yes
Short-Doyle	Dept Alcohol & Drug Abuse	No	Non-EDS	74	Yes
Inpatient Mental Health	DMH	No	EDS	63	Yes
Disproportionate Share Hospital Payments (credit)	DHS (Rates)	No	None	--	No
State Devlpmtl Ctrs (formerly State Hosp-DD)	DDS	Yes	Non-EDS	56	Yes
State Mental Hospitals	DMH	No	Non-EDS	57	Yes
Family PACT (aid code 8H)	DHS (FPACT)	No	EDS	--	Yes

Non-Medi-Cal

CCS/GHPP	DHS (CMS)	No	EDS-CCS 35-file	03	No
California Medical Services Program	DHS (CMSP)	Yes (CMSP version)	EDS-CMSP 35-file		No
CHDP	DHS (CMS)	No	EDS-CHDP (PM 160)		No
Healthy Families (AC 9H)	MRMIB	No	HF File - 35-file (Note: 7X = 7X; 00 = CCS; 9H = CTP); Dental LPC for 7X		No
Healthy Families -- (One-Month) Bridge Program (AC 7X)	DHS	No			No
Breast Cancer Early Detection Program	DHS (BCEDP)	No	35-file		Yes

* EDS = EDS 35-file; Non-EDS = Non-EDS 35-file

** Person formerly in the Model NF Waiver and the Nursing Facility Waiver were put in the two new waivers NF A/B and NF Subacute waivers. There is not a one-to-one relationship between these two waivers and the two ones in the chart.

Table A-2, Medi-Cal Eligibles Sources

Medi-Cal

Program	Agency Establishing Eligibility	Reported Eligibles (CIDCUM) (Yes or NO)	Eligibility File	Comments
Medi-Cal, Not Otherwise Specified	County Welfare Offices	Yes	Monthly Medi-Cal Eligibility File	Elig status = 001 - 499
Medi-Cal, SOC/Uncertified	County Welfare Offices	Yes	Monthly Medi-Cal Eligibility File	Elig status = 500 - 599
SSI/SSP	Federal Social Security Administration	Yes	Monthly Medi-Cal Eligibility File	Elig status = 001 - 499
Family PACT (Aid Code 8H)	Family PACT Providers	No	Family PACT Monthly Extract File	60 cumulative MOEs; eligible if days gt 0 for MOE
Medicare (AC's 8A, Qualified Disabled Working Individual (QDWI); 8C, Specified Low-Income Medicare Beneficiary (SLMB), 8D, Qualified Individual (1), and 8K, Qualified Individual (2)) (Monthly premium of ~$2.17 paid per bene)	Counties	No	Monthly Medi-Cal Eligibility File	Elig status = 600 - 699
Presumptive Eligibility (AC's 7F and 7G)	Providers	No	None	Eligibles counted from paid claims

Non-Medi-Cal Eligibles

CCS/GHPP	Counties	No	CCS/GHPP files	
CHDP	Providers	No	CHDP files	
California Medical Services Program	Counties	No	Monthly Medi-Cal Eligibility File	Elig status = 001 - 498
Healthy Families (AC 9H)	MRMIB	No	Monthly Medi-Cal Eligibility File	Elig status = 651, 652, 671, 672, 691, and 692.
Healthy Families -- Medi-Cal and CHDP Bridge to HF (AC 7X and 8X)	DHS	No	Monthly Medi-Cal Eligibility File	Elig status = 001 - 498
Breast Cancer Early Detection Program	DHS	No	BCEDP File	

APPENDIX B

DETAIL OF LAG FACTORS FOR ELIGIBILITY RETROACTIVITY

The following tables provides information on the rate at which retroactivity occurs for the Medi-Cal population. The methodology was to take the current (newly created) file for August 1998 and categorize the eligibles by aid code group (see below). Then the file created for September 1998 reporting August eligibles was used to aggregate these eligibles. Files subsequently created through January 1999 reporting on August eligibles were processed as well.

Group Designations

Group 1 = SSI/SSP
Group 2 = Long Term Care
Group 3 = AFDC - Cash Grant
Group 4 = Medi-Cal Only, Families
Group 5 = Medi-Cal Only, Aged, Blind, Disabled, no share of cost
Group 6 = Share of Cost
Group 7 = OBRA
Group 8 = Miscellaneous

Number of Eligibles

	Grp1	Grp2	Grp3	Grp4	Grp5	Grp6	Grp7	Grp8	Othr
Aug-98	1,052,448	63,435	2,332,438	825,186	109,571	228	174,903	8,740	38
Sep-98	1,066,299	66,191	2,376,712	938,498	112,838	18,696	197,036	7,926	5,974
Oct-98	1,075,481	67,839	2,397,844	986,057	114,022	23,807	207,870	7,929	6,992
Nov-98	1,081,357	68,893	2,404,095	1,008,662	114,525	25,601	213,140	7,895	7,340
Dec-98	1,086,501	69,387	2,405,125	1,017,351	114,700	26,903	215,317	7,839	7,456
Jan-99	1,089,732	69,686	2,405,273	1,021,180	114,810	27,739	216,427	7,804	7,518

Percent Conversions

	Grp1	Grp2	Grp3	Grp4	Grp5	Grp6	Grp7	Grp8*	Othr
Aug-98	96.58	91.03	96.97	80.81	95.44	0.82	80.81	111.99	0.51
Sep-98	97.85	94.98	98.81	91.90	98.28	67.40	91.04	101.56	79.46
Oct-98	98.69	97.35	99.69	96.56	99.31	85.83	96.05	101.60	93.00
Nov-98	99.23	98.86	99.95	98.77	99.75	92.29	98.48	101.17	97.63
Dec-98	99.70	99.57	99.99	99.63	99.90	96.99	99.49	100.45	99.18
Jan-99	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

* Note that the greater-than100% figure probably represents eligibles who were either switched to another aid code grouping, or were found ineligible during this retroactive period.

Aid Codes by Group

Group 1	20, 22, 60, 62, 10, 12, 28, 68, 18
Group 2	53, 63, 23, 13
Group 3	06, 30, 32, 33, 35, 38, 40, 42, 43, 77, 78, 3A, 3C, 3P, 3R, 3G, 3H, 4C
Group 4	72, 74, 7A, 7C, 07, 47, 69, 79, 86, 45, 4K, 04, 5K, 03, 82, 7M, 7P, 7R, 34, 44, 48, 49, 70, 75, 76, 7F, 7G, 39, 54, 59
Group 5	24, 64, 14, 25, 26, 6A, 46, 36, 66, 6C, 15, 16
Group 6	87, 83, 27, 37, 65, 67, 17
Group 7	55, 58, 5F
Group 8	51, 52, 56, 57, 81, 73, 80, 01, 0A, 02, 08, 71, 7H
Other =	All other aid codes

APPENDIX C

MISCELLANEOUS INFORMATION: FFS PAID CLAIMS FILE

Because the FFS Paid Claim files are used for so much of the claims research within the Medical Care Statistics Section, I offer here more detailed information about aspects of this file.

A. Medicare Crossover Claims

Some claims appearing on our claims files were first processed by one of the two Medicare fiscal intermediaries, then sent to EDS to adjudicate and pay any residual amount owed the provider. Because the claim has already been determined by Medicare to be a payable claim, EDS must process the claim whether the claim information is valid or not. Beneficiary data is, for example, not always valid, and thus should, whenever possible, be bypassed by researchers when processing claims data. Similiarly, claims from sources other than EDS and Delta Dental have data which should be used only warily.

We recently analyzed some of the monetary variables on crossover claims using a 5% sample of claims from calendar years 1995, 1997, and 2000. First, we found that the percent of adjustments were 0.7%, 1.7%, and 9.2% for the three years of 1995, 1997, and 2000, respectively. Next, we wanted to quantify how often crossover claims (non-adjustments) were paid at zero amount. This would be important to know if a researcher wanted to delete all crossovers from his/her analysis, but was not sure of the impact on the total paid amounts, especially by claim type. A break-out of this information is shown here.

Zero vs. Non-Zero Paid Amounts for Non-Adjustment Crossover Claims

		Drug	Inpatient	Medical Outpatient		Total
CY95	Non-zero Payments	30	10,508	125,099	22,389	158,026
	Zero Payment	9	8,486	15,588	22,643	46,726
	% zero payments of all	**23.1%**	**44.7%**	**11.1%**	**50.3%**	**22.8%**
CY97	Non-zero Payments	60	18,533	383,108	26,077	427,778
	Zero Payment	8	2,465	16,508	24,651	43,632
	% zero payments of all	**11.8%**	**11.7%**	**4.1%**	**48.6%**	**9.3%**
CY00	Non-zero Payments	258	13,960	148,840	32,534	195,592
	Zero Payment	65	6,428	29,018	24,291	59,802
	% zero payments of all	**20.1%**	**31.5%**	**16.3%**	**42.7%**	**23.4%**

If crossover claims were kept, but one wanted to know how to account for the paid amounts at the detail vs. the header level (Note: claim records are often provided researchers at the detail level for the non-inpatient claim type), it would be important to know any differences in aggregate between these two fields. An analysis from our 5% sample files provided the following:

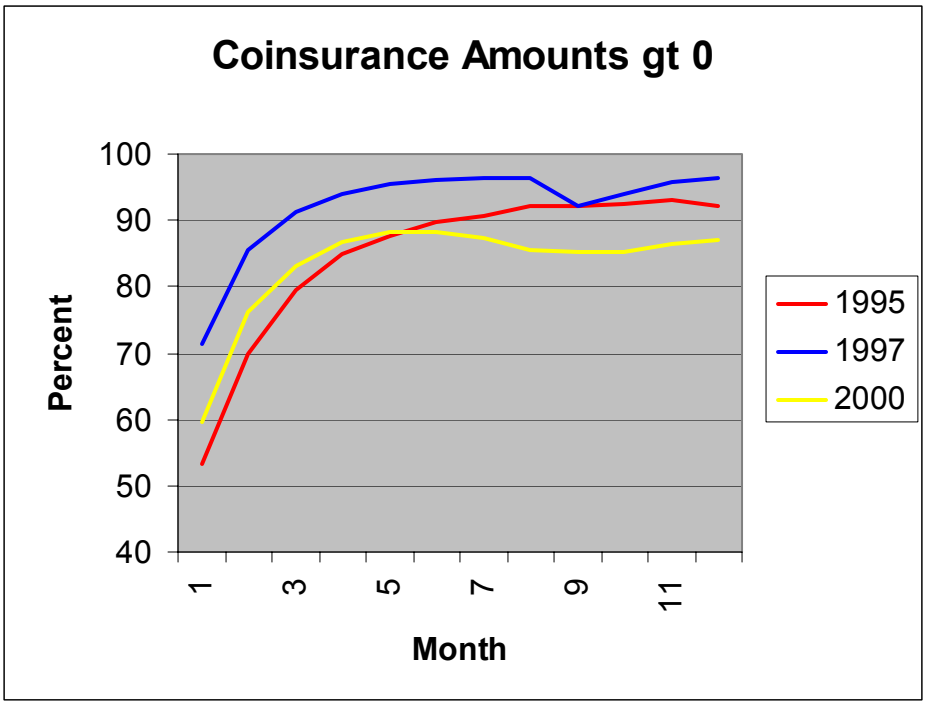
Aggregate paid amount of header vs. detail

		Header	Detail	% Header of detail
CY95	All	$11,492,488.08	$11,671,128.06	98.5%
	without Inpat	$ 6,744,162.78	$ 6,872,392.80	**98.1%**
CY97	All	$26,612,571.18	$27,459,715.40	96.9%
	without Inpat	$15,716,259.82	$15,888,039.86	**98.9%**
CY00	All	$20,053,524.08	$19,710,723.16	101.7%
	without Inpat	$10,520,071.77	$10,503,476.33	**100.2%**

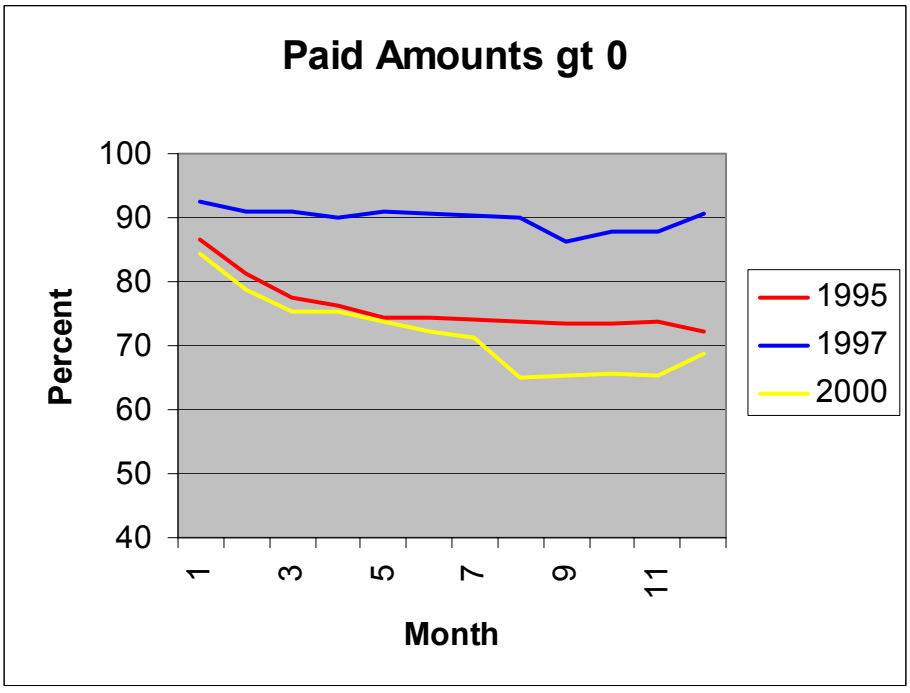
This indicates that the detail paid amounts generally match the header paid amount on each claim.

In over 98% of the cases, it was the first claim line which carried all the detail paid amount, and these almost always had a blank value in the procedure code variable.

Finally, we looked at the paid and coinsurance amount fields to see if there is any trend by calendar month of service insofar as the dollar amounts greater than zero. For header paid amounts, we derived the following:



Similarly, for the coinsurance amounts taken from the header we derived this chart:



B. Header vs. Segment Paid Amounts

In most cases, the paid amounts on the claim record segments (representing the individual claim lines) will add up to the paid amount on the header for that record. Occasionally, however, they will differ. In considering non-inpatient claims, the following are examples of this occurring.

1. Provider number prefix is ZZW

The paid amount on the segment is adjusted by a factor on file with the Department when the provider number prefix is ZZW (Los Angeles waiver outpatient facility), resulting in an actual paid amount different than the paid segment amount. These factors are normally less than 1.

2. Pharmacy claims

Dispensing and professional fees are added to the segment amounts, resulting in an actual paid amount higher than the paid segment amount. These fees usually amount to less than $10.

3. Patient/Liability/Share of Cost

When there is a patient liability/share of cost (SOC) amount, this reduces the paid segment amount by this SOC amount, except when the SOC is greater than the paid segment amount, in which case the resulting actual paid amount is zero (instead of negative). In the latter case, the SOC amount seems to be the same as the billed amount.

4. Third Party Liability

The above regarding patient liability/SOC seem to apply in a similar way to third party liability, except that this latter amount seems to be the same as the paid segment amount.

C. Beneficiaries in State Hospitals

Almost all services for beneficiaries confined to State Hospitals receive them through those hospitals. Occaisionally, these beneficiaries will have to receive care from other providers (presumedly because the State Hospitals cannot provide all medical care). However, the percent of those receiving outside is very small. When this does occur, if the outside provider is a Medi-Cal provider and the individual is Medi-Cal eligible, the provider is given Medi-Cal info and instructed to bill thru the

Fiscal Intermediary (most likely EDS.) If the provider is not a Medi-Cal provider and/or the individual is not Medi-Cal eligible, the facility pays the provider (facilities are instructed to search for Medi-Cal providers if individual is Medi-Cal eligible to minimize State Hospital expenditures.) The amount of facility paid services to outside providers is reported & included in the State Hospital cost reports utilized to calculate rates for the respective State Department to bill DHS. All state facility individuals billed to DHS in the monthly Medi-Cal billing cycle are Medi-Cal eligible. Approximately 96% of the DDS developmental center population are Medi-Cal eligible; less than 10% of the DMH state hospital population are Medi-Cal eligible.

D. Emergency Room Services – A Definition

Emergency services can be defined using the following criteria:

1. Emergency Indicator value of 'Y'; and,

2. Place of Service value of '0'; and,

3. Include all claims with a Vendor Code of 20, 22, 52, 62, 75, or 77, plus
has a Procedure Code of (99062 thru 99065, or 99281 thru 99285,
or Z7502, Z7504, Z7506, Z7508, Z7510, Z7610, or Z7612), plus
has a Place of Service value of '2' or '5'.

E. Long Term Care Eligibles - A Definition

Here are some guidelines to classify beneficiaries into long term care aid codes. For more specific criteria, it is recommended that the researcher cross-tabulate aid code by provider type 17 using the FFS Expenditures Summary Tables available on the MCSS website.

1. Exclude claims with vendor code 47, 56, 57, and 80 (this exclude State hospital claims also).

2. Include benes with an aid code of 13, 23, 53, and 63. Also, because there are eligibles not in one of these aid codes with an LTC claim, remove all benes from the eligible file found under 1 above.